An overview of importance splitting for rare event simulation

2010 Eur. J. Phys. 31 1295

(http://iopscience.iop.org/0143-0807/31/5/028)

IOP PUBLISHING EUROPEAN JOURNAL OF PHYSICS

Eur. J. Phys. **31** (2010) 1295–1303 doi:10.1088/0143-0807/31/5/028

An overview of importance splitting for rare event simulation

Jérôme Morio[1], **Rudy Pastel**[1] **and François Le Gland**[2]

[1] Office National d'Etudes et Recherches Aérospatiales, The French Aerospace Lab, Long-term Design and System Integration Department (ONERA-DPRS-SSD), BP72, 29 avenue de la Division Leclerc, FR-92322 Chatillon Cedex, France
[2] INRIA Rennes, ASPI Applications of Interacting Particle Systems to Statistics, Campus de Beaulieu 35042, Rennes, France

E-mail: jerome.morio@onera.fr

Received 8 June 2010, in final form 16 July 2010
Published 31 August 2010
Online at stacks.iop.org/EJP/31/1295

Abstract
Monte Carlo simulations are a classical tool to analyse physical systems. When unlikely events are to be simulated, the importance sampling technique is often used instead of Monte Carlo. Importance sampling has some drawbacks when the problem dimensionality is high or when the optimal importance sampling density is complex to obtain. In this paper, we focus on a quite novel but somehow confidential alternative to importance sampling called importance splitting.

1. Introduction

When one considers a physical black-box model, you have a certain number of input parameters and a few equations that use those inputs to give a set of outputs. This type of model is usually deterministic, meaning that you always get the same outputs for a given set of inputs, no matter how many times you re-calculate. Monte Carlo simulation is a method for iteratively evaluating a deterministic model using sets of random numbers as inputs. This method is often used when the model is nonlinear, complex or involves more than just a couple uncertain parameters. A simulation can typically require over 10 000 evaluations of the black-box model.

Monte Carlo simulation is categorized as a sampling method because the inputs are randomly generated from probability distributions to simulate the process of sampling from an actual population. It is useful in studying systems with a large number of coupled degrees of freedom, such as fluids, disordered materials, strongly coupled solids and cellular structures.

When unlikely events are to be simulated, the importance sampling (IS) technique is often used instead of Monte Carlo because of the computation burden involved in Monte Carlo simulations. In the world of physics (safety, finance, nuclear, defence, spatial, etc),

estimating rare event probability or low integral with a valuable accuracy has become very important. Indeed, questions like 'what is the probability of collision between a satellite and spatial debris ?' cannot be answered easily with the Monte Carlo approach. To reduce the variance on the probability estimator, the most well-known algorithm is IS which consists in generating random weighted samples from an auxiliary distribution rather than the distribution of interest. The main difficulty of IS is to determine a valuable auxiliary distribution which is not trivial in multidimensional cases. This technique has been notably treated in this journal in [1–3]. A recent alternative to IS is importance splitting (ISp). Instead of estimating one probability through a very tough simulation, one should consider the estimation of several conditional probabilities that are easier to evaluate by simulation. We show in this paper the great potential of this technique in a very general case of probability estimation. This paper is intended for graduate students and specialists in the field of probability or integral estimation in physics.

2. Usual methods to estimate a probability

2.1. Monte Carlo methods

Let us consider a d-dimensional random variable X with a probability density function (PDF) f_0 and estimate the probability that $P(\phi(X) > S)$ with ϕ, a continuous scalar function $\phi : \mathbb{R}^d \to \mathbb{R}$ and S a threshold. A simple way to estimate this probability is to consider Monte Carlo methods [4–8]. For that purpose, one generates independent and identically distributed samples $X_1, \ldots, X_N$ from the PDF f_0 and then estimates the probability with

$$P^{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(X_i)>S} \tag{1}$$

where $\mathbf{1}_{\phi(X_i)>S}$ is equal to 1 if $\phi(X_i) > S$ and 0 otherwise. The relative deviation of the estimator P^{MC} is given by the ratio $\frac{\sigma_{P^{\mathrm{MC}}}}{P^{\mathrm{MC}}}$ with $\sigma_{P^{\mathrm{MC}}}$, the standard deviation of P^{MC}. Knowing the true probability P that $(\phi(X) > S)$, one has

$$\frac{\sigma_{P^{\mathrm{MC}}}}{P^{\mathrm{MC}}} = \frac{1}{\sqrt{N}} \frac{\sqrt{P - P^2}}{P}. \tag{2}$$

Considering rare event probability estimation, that is when P takes low values, one has

$$\lim_{P\to 0} \frac{\sigma_{P^{\mathrm{MC}}}}{P^{\mathrm{MC}}} = \lim_{P\to 0} \frac{1}{\sqrt{NP}} = +\infty. \tag{3}$$

The relative deviation of Monte Carlo estimation is very important and thus one can conclude that Monte Carlo methods are not adapted to rare event probability estimation.

2.2. Importance sampling

One possible remedy for Monte Carlo failure is to consider the IS method [1–3, 9–13]. The objective of IS is to reduce the estimation variance without increasing the number of samples N. The idea is to generate samples $X_1, \ldots, X_N$ from an auxiliary PDF h and then estimate P in the following way:

$$P^{IS} = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i) \mathbf{1}_{\phi(x_i)>S} \frac{f_0(x_i)}{h(x_i)}. \tag{4}$$

The variance reduction depends on the choice of the auxiliary PDF h. If h is well chosen, the estimation results are drastically better than Monte Carlo methods. If h is not adapted to

the problem, IS can lead to worse results than Monte Carlo methods. In theory, an optimal auxiliary PDF h_{opt} can be obtained, but it is not available in practice. This PDF h_{opt} is optimal since it enables us to minimize the variance estimation but its computation requires the knowledge of the probability that has to be estimated. The main issue of IS is thus the optimization of the auxiliary PDF h to approach this PDF h_{opt}. This optimization is not trivial and several methods (parametric [12] or not [11], adaptive [11] or not [14]) can be useful but are not always efficient. The multidimensional case, that is when X follows a multidimensional PDF, is often problematic because of the curse of dimensionality.

3. Importance splitting

In this paper, we consider a very efficient but not well-known algorithm called ISp [15–18]. ISp is an alternative to Monte Carlo methods and to IS. We first propose to present the principle of ISp and to apply it on a simple test case.

3.1. Principle

Considering the set $\mathbf{A} = \{x \in \mathbb{R}^d | \phi(x) > S\}$, the objective of this paper is to determine the density in this set A since one has $P(X \in \mathbf{A}) = P(\phi(X) > S)$. The principle of ISp is to iteratively estimate supersets of the set A and then to estimate $P(X \in \mathbf{A})$ with conditional probabilities.

Let us define $\mathbf{A}_0 = \mathbb{R}^d \supset \mathbf{A}_1 \supset \cdots \supset \mathbf{A}_{n-1} \supset \mathbf{A}_n = \mathbf{A}$, a decreasing sequence of $\mathbb{R}^d$ subsets with smallest element $\mathbf{A} = \mathbf{A}_n$. The probability $P(X \in \mathbf{A})$ can then be rewritten in the following way:

$$P(X \in \mathbf{A}) = \prod_{k=1}^{n} P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$$

where $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$ is the probability that $X \in \mathbf{A}_k$ knowing that $X \in \mathbf{A}_{k-1}$. An optimal choice of the sequence $\mathbf{A}_k$, $k = 0, \ldots, n$, is given when $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1}) = p$, where p is a constant, that is when all the conditional probabilities are equal. The variance of $P(X \in \mathbf{A})$ is indeed minimized in this configuration as shown in [19, 20]. Consequently, if each $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1})$ is well estimated, then the probability $P(X \in \mathbf{A})$ is estimated more accurately with ISp than with a direct estimation by Monte Carlo.

3.2. Defining the A_k sequence

The subset $\mathbf{A}_k$ sequence is easily evaluated in the following way. Indeed, it can be defined with $\mathbf{A_k} = \{X \in \mathbb{R}^d | (\phi(x)) > S_k\}$ for $k = 1, \ldots, n$ with $S = S_n > S_{n-1} > \cdots > S_k > \cdots > S_1$. Determining the sequence $\mathbf{A}_k$ is equivalent to choosing some values for S_k, with $k = 1, \ldots, n$.

If one has sufficient knowledge on the density of $\phi(X)$, it is possible to set a priori the values of S_k for $k = 1, \ldots, n$. Nevertheless, it is in general not the case and this choice of thresholds is hardly ever optimal.

The values of S_k for $k = 1, \ldots, n$ can also be obtained in an adaptive manner. Let us define f_k, the density of X restricted to the set $\mathbf{A}_k$ and μ_k, the density of $\phi(X)$ when X is restricted to the set $\mathbf{A}_k$. Generate $X_1^{(k)}, \ldots, X_{N_k}^{(k)}$, N_k samples from the density f_k. The Monte Carlo estimator of the μ_k cumulative distribution function G_{μ_k} is given by

$$G_{\mu_k}(y) = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{1}_{\phi(X_i^{(k)}) \leqslant y}. \tag{5}$$

The Monte Carlo estimator of the α-quantile of the density μ_k is given by

$$q_\alpha^{(k)} = \inf\{y, G_{\mu_k}(y) \geqslant \alpha\}. \tag{6}$$

Set $S_{k+1} = q_\alpha^{(k)}$. The subset $\mathbf{A}_{k+1}$ is then defined with $\mathbf{A_{k+1}} = \{X \in \mathbb{R}^d | (\phi(x)) > S_{k+1} = q_\alpha^{(k)}\}$. Let us then estimate $P(X \in \mathbf{A}_{k+1} | X \in \mathbf{A}_k)$:

$$E(P(X \in \mathbf{A}_{k+1} | X \in \mathbf{A}_k)) = E_{\mu_k}(P(X \in \mathbf{A}_{k+1})) = E_{\mu_k}(P(\phi(X) > S_{k+1})). \tag{7}$$

By definition of S_{k+1}, one has then

$$E\left(P(X \in \mathbf{A}_{k+1} | X \in \mathbf{A}_k)\right) = 1 - \alpha. \tag{8}$$

With this adaptive definition of S_{k+1}, an optimal sequence of A_k is then determined in the sense that all the $P(X \in \mathbf{A}_k | X \in \mathbf{A}_{k-1}) = 1 - \alpha$ are a constant [19, 20]. The number n of the subset depends on the value of P and α since $P = (1 - \alpha)^{n-1} * P(X \in \mathbf{A}_n | X \in \mathbf{A}_{n-1})$. The term n is not known at the start of the algorithm. In order to limit the computation time, the quantile parameter α has to be adjusted to prevent a too large n.

3.3. Generating samples from the density f_k

Unfortunately, generating directly independent samples from the f_k conditional densities in most cases is impossible as they are usually unknown [21]. Nevertheless, ISp provides an iterative way to do it, yet in a dependent fashion. A Markovian kernel can be seen here as a collection of probability density functions indexed on x : $\forall\, x \in \mathbb{R}^d$, $M(x, \cdot)$ is a PDF. It is assumed that we have an f_0-reversible Markovian kernel $M(x, \cdot)$ such that the reversibility equation holds

$$\forall\, x \in \mathbb{R}^d, \quad \forall\, y \in \mathbb{R}^d, \quad f_0(x)M(x, y) = f_0(y)M(y, x).$$

This reversibility equation can be seen as a detailed balance statement. Moreover, the above reversibility equation embeds two results.

(i) If you generate the random variable Y with density $M(X, \cdot)$, where $X \sim f_0$, Y is distributed according to f_0 as well. The demonstration is given in the following. Let us define g as an integrable function of $\mathbb{R}^d$. The expected value E of $g(Y)$ is given by

$$E(g(Y)) = \int_{\mathbb{R}}^d f_0 M(y, .) g(y)\, \mathrm{d}y = \int_{\mathbb{R}}^d \int_{\mathbb{R}}^d f_0(x) M(x, y) g(y)\, \mathrm{d}x\, \mathrm{d}y. \tag{9}$$

Using the reversability equation, one obtains

$$E(g(Y)) = \int_{\mathbb{R}}^d \int_{\mathbb{R}}^d f_0(y) M(y, x) g(y)\, \mathrm{d}x\, \mathrm{d}y = \int_{\mathbb{R}}^d f_0(y) \left(\int_{\mathbb{R}}^d M(y, x)\, \mathrm{d}x \right) g(y)\, \mathrm{d}y. \tag{10}$$

We can finally show that

$$E\left(g(Y)\right) = \int_{\mathbb{R}}^d f_0(y) g(y)\, \mathrm{d}y = \int_{\mathbb{R}}^d f_0(x) g(x)\, \mathrm{d}x = E(g(X)). \tag{11}$$

This equality is true for any integrable function of $\mathbb{R}^d$. The random variables X and Y follow thus the same PDF.

(ii) If you take $(X^1, \ldots, X^N)$ a sample set such that $\forall\, i \in \{1, \ldots, N\}$, $X^i \sim f_0$ to generate $(Y^1, \ldots, Y^N)$ such that $\forall\, i \in \{1, \ldots, N\}$, $Y_i \sim M(X^i, \cdot)$, then both sets have the same statistical properties and it is impossible to know through statistics which one generated the other.

With such a kernel and $X_k \sim f_k$, one can distribute random variable Ξ_k according to f_k:

$$\Xi_k = \Xi_k(X_k) = \begin{cases} M(X^k, \cdot) & \text{if} \quad M(X^k, \cdot) \in A_k \\ X^k & \text{otherwise} \end{cases}.$$

This proposal/refusal algorithm enables us to generate any number of samples according to f_k in a relatively simple manner. It also enables us to keep constant the number of samples to estimate each $P(X \in \mathbf{A}_{k+1} | X \in \mathbf{A}_k)$. This operation has to be applied for each density f_k, $k = 1, \ldots, n$. The generated set holds *no independent* variables identically distributed according to f_{k+1}. Up to now, there is no way to do this in a independent fashion. However, under mild conditions, it can be shown that applying the proposal/refusal method several times may decrease variance. This process can be repeated iteratively to generate according to the other conditional densities.

The fundamental question is therefore how to define such a kernel for a density X. The most general and arduous way is to solve the reversibility equation but this is in general not possible. For that purpose, the Metropolis–Hastings algorithm enables us to simulate a Markovian reversible kernel in the general case [21]. Nevertheless, some results exist when the density X is a Gaussian multivariate PDF. Indeed, if X is a centred Gaussian, a Markovian reversible kernel is

$$X \sim \mathcal{N}(0_d, I_d) \qquad \text{and} \qquad M(X, \cdot) \sim \frac{X + c\mathcal{N}(0_d, I_d)}{\sqrt{1 + c^2}}.$$

It is a valuable theoretical reversible kernel for the Gaussian distribution. The choice of an appropriate value for the kernel parameter c is not obvious and is still an open question. In some advanced algorithm, c is determined in an adaptive manner.

All the required steps for ISp have been defined in the previous sections. Let us apply this technique on general cases to compare the results of the techniques with Monte Carlo simulations.

3.4. Implementation

We propose to analyse the different stages of the algorithm ISp to estimate $P(\phi(X) > S) = P(X \in \mathbf{A})$.

(1) Set $k = 0$.
(2) Generate N_k samples $X_1^{(k)}, \ldots, X_{N_k}^{(k)}$ from $f_k(X)$.
(3) Estimate the α-quantile $q_\alpha^{(k)}$ of the samples $X_1^{(k)}, \ldots, X_{N_k}^{(k)}$.
(4) Determine the subset A_{k+1} with $\mathbf{A_{k+1}} = \left\{X \in \mathbb{R}^d | (\phi(x)) > q_\alpha^{(k)}\right\}$ and the conditional density f_k.
(5) If $q_\alpha^{(k)} < S$, set $k = k + 1$ and go back to stage (2) of the algorithm. Otherwise, set $k = k + 1$ and estimate the probability with

$$P^{\text{ISp}} = (1 - \alpha)^k \times \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{1}_{\phi(X_i^{(k)}) > S}.$$

4. Application of importance splitting

In this section, we propose to evaluate the performance of the ISp algorithm on different test cases.

```
N=10000; % number of samples generated for each intermediate conditional estimation
alpha=0.8; %  quantile level for adaptive threshold positioning
X=randn(N,1); % generation of initial samples
c=1; % parameter of markovian kernel
S=3; % threshold to be exceeded.
q_alpha=quantile(X,alpha); % estimation of q-quantile
i=0;
while (q_alpha<S)
w=(X>q_alpha);
Y=randsample(X,N,'true',w); % resampling
p=(Y+c*randn(N,1))/sqrt(1+c^2);% markovian kernel application.
X=(p>=q_alpha).*p+(p<q_alpha).*Y; % new population of particules
q_alpha=quantile(X,alpha);% position of the next threshold
i=i+1;
end

proba=(1-alpha)^i*(mean(X>S)); % probability estimation with importance splitting
```

Figure 1. Matlab code of importance splitting.
(This figure is in colour only in the electronic version)

Table 1. Estimation results $P(Y > S)$ with ISp for $\alpha = 0.5$.

N	S	$P^{\mathrm{ISp}}(Y > S)$	Relative error (%)	Theoretical probability $P(Y > S)$
100 000	3	1.34×10^{-3}	5	1.350×10^{-3}
160 000	4	3.17×10^{-5}	9.9	3.16×10^{-5}
220 000	5	2.91×10^{-7}	14.9	2.87×10^{-7}
310 000	6	1.40×10^{-9}	45	9.87×10^{-10}

4.1. Simple Gaussian case

We first consider a simple case where X is a one-dimensional Gaussian PDF, that is, $\mathcal{N}(0, 1)$ and ϕ is equal to I_1, that is $Y = \phi(X) = X$. Let us estimate the probability that $P(Y > S)$ for different values of S. For that purpose, we use the ISp code proposed in figure 1. The number of samples N_k is set to 10 000 and the kernel parameter c is arbitrarily chosen equal to 1. The results obtained with this code are given in table 1 for $\alpha = 0.5$ and figure 2 presents the iterative evolution of the parameters $q_\alpha^{(i)}$ and of the conditional densities f_i. In tables 2, the corresponding Monte Carlo results are also presented and can be compared to those obtained with ISp. In tables 3 and 4, the ISp estimation results are given for different values of α. ISp accurately estimates the probability since the relative error is lower than with Monte Carlo and the probability estimate is closer to the theoretical probability. The proposed results show the ISp ability to be a valuable rare event probability estimation method.

4.2. Multidimensional test case

In this last section, we consider the case where $X = (x_1, x_2, \ldots, x_5)$ follows a multidimensional Gaussian PDF with mean $(0, 0, 0, 0, 0, 0)$ and a covariance matrix equal

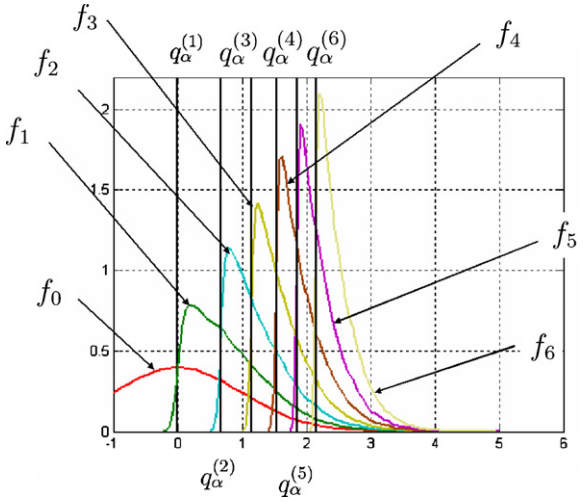

Figure 2. Evolution of the parameters $q_\alpha^{(i)}$ and of the conditional densities f_i estimated with a kernel density estimator on the simple Gaussian case for $i = 1, \ldots, 6$.

Table 2. Estimation results $P(Y > S)$ with the Monte Carlo method.

N	S	$P^{\mathrm{MC}}(Y > S)$	Relative error (%)	Theoretical probability $P(Y > S)$
100 000	3	1.36×10^{-3}	9	1.350×10^{-3}
160 000	4	2.95×10^{-5}	47	3.16×10^{-5}
220 000	5	4.54×10^{-7}	333	2.87×10^{-7}
310 000	6	0	?	9.87×10^{-10}

Table 3. Estimation results $P(Y > S)$ with ISp for $\alpha = 0.2$.

N	S	$P^{\mathrm{ISp}}(Y > S)$	Relative error (%)	Theoretical probability $P(Y > S)$
300 000	3	1.35×10^{-3}	3.8	1.350×10^{-3}
460 000	4	3.21×10^{-5}	6.8	3.16×10^{-5}
670 000	5	3.00×10^{-7}	12.1	2.87×10^{-7}
940 000	6	1.32×10^{-9}	21.7	9.87×10^{-10}

to I_5, the identity matrix of $\mathbb{R}^5$. The function ϕ is given by the following function, called the Ackley function:

$$Y = \phi(X) = -20 \exp\left(-0.2 \sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d} \cos(2\pi x_i)\right) + 20 + e. \tag{12}$$

The Ackley function is a well-known scalar function with multidimensional inputs which is often used as a test function for the global optimization algorithm. This case is interesting to deal with since multidimensional sampling is often tricky. It is indeed one of the limitations of IS since the IS optimal auxiliary distribution is expensive computationally.

Table 4. Estimation results $P(Y > S)$ with ISp for $\alpha = 0.8$.

N	S	$P^{\text{ISp}}(Y > S)$	Relative error (%)	Theoretical probability $P(Y > S)$
40 000	3	1.36×10^{-3}	7.6	1.350×10^{-3}
60 000	4	3.26×10^{-5}	13.8	3.16×10^{-5}
90 000	5	3.14×10^{-7}	30.3	2.87×10^{-7}
120 000	6	1.28×10^{-9}	68.4	9.87×10^{-10}

Table 5. Estimation results $P(\phi(X) > S)$ with ISp for $\alpha = 0.5$.

N	S	$P^{\text{ISp}}(Y > S)$	Relative error (%)
70 000	8	3.97×10^{-3}	4.5
120 000	9	1.58×10^{-4}	9.7
180 000	10	2.31×10^{-6}	22
250 000	11	1.75×10^{-8}	60
320 000	12	1.89×10^{-10}	200

Table 6. Estimation results $P(\phi(X) > S)$ with Monte Carlo.

N	S	$P^{\text{MC}}(Y > S)$	Relative error (%)
70 000	8	3.93×10^{-3}	5.3
120 000	9	1.54×10^{-4}	24
180 000	10	2.55×10^{-6}	149
250 000	11	0	?
320 000	12	0	?

Table 7. Estimation results $P(\phi(X) > S)$ with ISp for $\alpha = 0.2$.

N	S	$P^{\text{ISp}}(Y > S)$	Relative error (%)
240 000	8	3.99×10^{-3}	4.8
390 000	9	1.60×10^{-4}	6.9
580 000	10	2.35×10^{-6}	14.9
790 000	11	2.02×10^{-8}	45.2
1 020 000	12	5.10×10^{-10}	80.5

We apply the ISp algorithm on this test case with the same parameters as in the previous section. The probability estimation results with ISp are given in table 5. For comparison purpose, we have estimated with the Monte Carlo method these probabilities in table 6. The same number of evaluations of the function ϕ is made with both methods. ISp shows great improvement for rare event simulation and estimation since the relative error of probability estimation is reduced when compared to Monte Carlo. In tables 7 and 8, the ISp estimation results are given for different values of α.

Table 8. Estimation results $P(\phi(X) > S)$ with ISp for $\alpha = 0.8$.

N	S	$P^{\mathrm{ISp}}(Y > S)$	Relative error (%)
30 000	8	3.99×10^{-3}	6.8
50 000	9	1.54×10^{-4}	15.2
70 000	10	2.27×10^{-6}	35.0
120 000	11	1.67×10^{-8}	110
140 000	12	1.25×10^{-10}	265

5. Conclusion

In this paper, we present a quite novel approach to estimate rare events called ISp. This algorithm is an alternative to the well-known IS and we showed on simple cases that it is well adapted to the estimation of rare events. An example of the ISp Matlab code is provided in this paper. We show that it can be used with simplicity although a field of expertise is necessary to optimize the different parameters of the algorithm.

References

[1] Borcherds P H 2000 Importance sampling: an illustrative introduction *Eur. J. Phys.* **21** 405–11
[2] Denny M 2001 Introduction to importance sampling in rare-event simulations *Eur. J. Phys.* **22** 403–11
[3] Morio J 2010 How to approach the importance sampling density *Eur. J. Phys.* **31** 41–8
[4] Silverman B 1986 *Density Estimation for Statistcs and Data Analysis* (London: Chapman and Hall)
[5] Mikhailov G A 1999 *Parametric Estimates by the Monte Carlo Method* (Utrecht NED: VSP)
[6] Sobol I M 1994 *A Primer for the Monte Carlo Method* (Boca Raton, FL: CRC Press)
[7] Robert C P and Casella G 2005 *Monte Carlo Statistical Methods* (New York: Springer)
[8] Niederreiter H and Spanier J 2000 *Monte Carlo and Quasi-Monte Carlo Methods* (Berlin: Springer)
[9] Beichl I and Sullivan F 1999 The importance of importance sampling *Comput. Sci. Eng.* **1** 71–3
[10] Botev Z I, Kroese D and Taimre T 2007 Generalized cross-entropy methods with applications to rare-event simulation and optimization *Simulation* **11** 785–806
[11] Zhang P 1996 Nonparametric importance sampling *J. Am. Stat. Assoc.* **91** 1245–53
[12] Rubinstein R 1997 Optimization of computer simulation models with rare events *Eur. J. Oper. Res.* **99** 89–112
[13] Glynn P 1996 Importance sampling for Monte Carlo estimation of quantiles *Mathematical Methods in Stochastic Simulation and Experimental Design: Proc. 2nd St. Petersburg Workshop on Simulation* (Publishing House of Saint Petersburg University) pp 180–5
[14] Bucklew J A 1990 *Large Deviation Techniques in Decision, Simulation, and Estimation* (New York: Wiley)
[15] Cerou F, Del Moral P, Furon T and Guyader A 2008 Rare event simulation for a static distribution *Proc. 7th Int. Work. on Rare Event Simulation (Rennes, France)* pp 107–15
[16] Cerou F and Guyader A 2007 Adaptive particle techniques and rare event estimation *ESAIM Proc. 19* pp 65–72
[17] L'Ecuyer P, Demers V and Tuffin B 2006 Splitting for rare event simulation *Proc. 2006 Winter Simulation Conf.* pp 137–48
[18] Glasserman P, Heidelberger P, Shahabuddin P and Zajic T 1996 Splitting for rare event simulation: analysis of simple cases *Proc. 1996 Winter Simulation Conf.* pp 302–8
[19] Lagnoux A 2006 Rare event simulation *Probab. Eng. Informational Sci.* **20** 45–66
[20] Cerou F, Del Moral P, Le Gland F and Lezaud P 2006 Genetic genealogical models in rare event analysis *INRIA Report 1797* pp 1–30
[21] Tierney L 1994 Markov chains for exploring posterior distributions *Ann. Stat.* **22** 1701–62